UNITED STATES DEPARTMENT OF COMMERCE
Economics and Statistics Administration
U.S. Census Bureau
Washington, DC 20233-0001

October 25, 2007

Final
DSMD
Signed

MEMORANDUM FOR Walter C. Odom, Jr.
Division Chief, Administrative and Customer Services Division

From: Ruth Ann Killion
Division Chief, Demographic Statistical Methods Division

Subject: Survey of Income and Program Participation (SIPP) 2004 Panel: Source and Accuracy Statement for Longitudinal Analysis of Waves 1 to 4 Public Use Files (S&A-5)[1]

Attached is the Source and Accuracy Statement for longitudinal analysis of the first 4 waves of the 2004 Survey of Income and Program Participation (SIPP).

Attachment

cc:
A. Shields (ACSD)
C. Landman (DSD)
T. Blatt
N. McKee
Z. McBride
P. Benton
D. Daily
M. Thrift (DID)
D. Johnson (HHES)
J. Day
J. Eargle
R. Kominski
C. Nelson
K. Bauman
J. Fields
J. Hisnanick
M. O'Connell
T. Palumbo
P. Salopek
S. Stern
P. Flanagan (DSMD)
SIPPB

DSMD:\ADC-CPHISD\SIPPB\Final Memos\S&A\S&A04_Long(S&A-5).wpd

[1]This source and accuracy statement can also be accessed through the U.S. Census

Bureau website at “*http://www.sipp.census.gov/sipp/source*.html”

SOURCE AND ACCURACY STATEMENT FOR THE SURVEY OF INCOME AND PROGRAM PARTICIPATION (SIPP) 2004, 12-WAVE LONGITUDINAL FILE[1]

DATA COLLECTION AND ESTIMATION

Source of Data. The data were collected in the 2004 Panel of the Survey of Income and Program Participation (SIPP). The population represented in the 2004 SIPP (the population universe) is the civilian noninstitutionalized population living in the United States. The institutionalized population, which is excluded from the universe, is composed primarily of the population in correctional institutions and nursing homes (91 percent of the 4.1 million institutionalized people in Census 2000).

The 2004 Panel of the SIPP sample is located in 351 Primary Sampling Units (PSUs), each consisting of a county or a group of contiguous counties. Of these 351 PSUs, 123 are self representing (SR) and 228 are non-self representing (NSR). SR PSUs have a probability of selection of one, NSR PSUs have a probability of selection of less than one. Within PSUs, housing units (HUs) were systematically selected from the master address file (MAF) used for the 2000 decennial census. To account for HUs built within each of the sample areas after the 2000 census, a sample containing clusters of four HUs was drawn from permits issued for construction of residential HUs up until shortly before the beginning of the panel. In jurisdictions that don’t issue building permits or that have incomplete addresses, we systematically sampled expected clusters of four HUs which were then listed by field personnel.

Sample households within a given panel are divided into four random subsamples of nearly equal size. These subsamples are called rotation groups and one rotation group is interviewed each month. Each household in the sample was scheduled to be interviewed at 4-month intervals over a period of roughly 4 years beginning in February 2004. The reference period for the questions is the 4-month period preceding the interview month. The most recent month is designated reference 4, the earliest month is reference month 1. In general, one cycle of four interviews covering the entire sample, using the same questionnaire, is called a wave. For example, Wave 1 rotation group 1 of the 2004 Panel was interviewed in February 2004 and data for the reference months October 2003 through January 2004 were collected.

The period covered by the 2004 12-Wave longitudinal file consists of 48 interview months (12 interviews) conducted from February 2004 to January 2008. Data for up to 51 reference months are available for persons on the file. Specific months available depend on the person’s rotation group and his/her sample entry or exit date. Also note that the availability of data on household composition begins with the first interview month of a rotation group.

---

[1]For questions or further assistance with the information provided in this document contact: Tracy Mattingly of the Demographic Statistical Methods Division on 301/763-6445 or via the email at Tracy.L.Mattingly@census.gov.

In Wave 1, we fielded a sample of about 62,700 HUs. About 11,300 of these HUs were found to be vacant, demolished, converted to nonresidential use, or otherwise ineligible for the survey. Interviews were obtained for about 43,700 of the eligible HUs. We did not interview approximately 7,700 eligible HUs in the panel because the occupants: (1) refused to be interviewed, (2) could not be found at home, (3) were temporarily absent, or (4) were otherwise unavailable. Thus, occupants of about 85 percent of all eligible HUs participated in the first interview of the panel.

For subsequent interviews, only original sample people (those in Wave 1 sample households and interviewed in Wave 1) and people living with them were eligible to be interviewed.
We will follow original sample people if they moved to a new address, unless the new address is more than 100 miles from a SIPP sample area. Then, we will attempt telephone interviews.

For the Panel, CY01, CY02 and CY03 weighting procedures, a person was classified as interviewed or noninterviewed based on the following definitions. (NOTE: A person may be classified differently for calculating different weights.) Interviewed sample persons (including children) were defined to be:

1) those for whom self, proxy, or imputed responses were obtained for each month of the appropriate longitudinal period, or

2) those for whom self or proxy responses were obtained for the first month of the appropriate longitudinal period and self, proxy, or imputed responses exist for each subsequent month until they were known to have died or moved to an ineligible address (foreign HUs, institutions, or military barracks).

The months for which persons were deceased or residing in an ineligible address were identified on the file. Noninterviewed persons were defined to be those for whom neither self nor proxy responses were obtained for one or more months of the appropriate longitudinal period (excluding imputed persons and persons who died or moved to an ineligible address).

It is estimated that roughly *161,100*[2] people were initially designated in the sample[3]. Approximately *112,300* people were interviewed in Wave 1; however, we did not interview approximately *19,700* of the sample persons in the panel because the occupants, (1) refused to be interviewed, (2) could not be found at home, (3) were temporarily absent, or (4) were otherwise unavailable. Thus, occupants of about 70 percent of all people initially designated in the sample participated in the first interview of the panel.
For the CY01 and CY02 weighting procedures, the eligible sample is considered to be all people interviewed in Wave 1 together with those joining the sample households at later times during

[2]All values given in italics in this paragraph are estimates.

[3]This approximation represents the number of HUs fielded in Wave1 multiplied by the average number of persons per household in Wave 1.

the panel. For the CY03, CY04 and 12-Wave (Panel) weighting procedures, a 53% sample cut occurring after Wave 8 further reduced the eligible sample. The CY01 weighting procedure classified about 87,600 people as interviewed and had a person nonresponse rate of 20.3%.

Estimation: The SIPP program produces weights for both cross-sectional and longitudinal analysis. What follows is an overview of the longitudinal estimation.
All people classified as interviewed for the longitudinal period (i.e., Panel, CY01, CY02, CY03 and CY04) are assigned positive weights for that period, while those classified as non-interviewed are assigned zero weights. Longitudinal weights are produced at the completion of Waves 4, 7 10 and 12. Newer panel weights covering longer time periods replace earlier panel weights until the end of the panel. We assigned an initial weight equal to the Wave 1 Cross Sectional household non-interview adjusted weight:

InitWt* = *HH Base Weight* * *Duplication Control Factor* * *Household Noninterview Factor,

where the HH Base Weight is equal to the inverse of the probability of selection of a person's household, the Household Noninterview Factor is applied to account for households which were eligible for the sample but which field representatives could not interview in Wave 1 and the Duplication Factor was applied to adjust for subsampling done in the field when the number of sampling units was much larger than expected.

Next we applied a Person Level Non-Interview Adjustment Factor (F_{NI}) to account for persons eligible for the sample but were classified as non-interviews due to their leaving the sample household, refusal to interview or other reasons. The factors were calculated individually for sample persons in each of 149 non-interview adjustment cells.

The last adjustment is the Second Stage Adjustment Factor, (F_{2S}). This incorporates an iterative adjustment process to population controls for selected demographic groups, performed to reduce the mean square error of the survey estimates. The control dates for the Panel, CY01, CY02, CY03 and CY04 are March 1, 2004, January 1, 2004, January 1, 2005, January 1, 2006 and January 1, 2007.

The final longitudinal weight is then:

$$FW_L = InitWt * F_{NI, J} * F_{2s}$$

Population Controls:
The survey's estimation procedure adjusts weighted sample results to agree with independently derived population estimates of the civilian noninstitutional population. This attempts to correct for undercoverage and thereby reduce the mean square error of the estimates. The national and state level population controls are obtained directly from the Population Division and are prepared each month to agree with the most current set of population estimates released by the Census Bureau's population estimates and projections program.

The national level controls are distributed by demographic characteristics as follows:

- Age, Sex, and Race (White Alone, Black Alone, and all other groups combined)
- Age, Sex, and Hispanic Origin

The state level controls are distributed by demographic characteristics as follows:

- State by Age and Sex
- State by Hispanic origin
- State by Race (Black Alone, all other groups combined)

The estimates begin with the latest decennial census as the base and incorporate the latest available information on births and deaths along with the latest estimates of net international migration.

Use of Person Weights. Panel weights are computed for persons who are in sample at Wave 1 and for whom data are obtained (either reported or imputed) for every month of the panel for which they were in scope for the survey. Calendar year weights are computed for people who had an interview covering the control date and for whom data are obtained (either reported or imputed) for every month of the calendar year for which they were in scope for the survey. The panel weight can be used to form monthly, quarterly, annual, or multi-year estimates for calendar years 2004 through 2007. The calendar year weight can be used to form monthly, quarterly, or annual estimates within a specific calendar year.

> **Example**, using the panel weight, one can estimate the number of people receiving TANF from January 2004 to January 2008. Using the CY03 weight, one can estimate the number of people receiving TANF for the third quarter of 2006.

Users should be forewarned to apply the appropriate weights given on weighting files before attempting to calculate estimates. The weights vary between units due to weighting adjustments, and following movers. If analysis is done for the general population without applying the appropriate weights, the results will be erroneous.

All estimates may be divided into two broad categories: longitudinal and cross-sectional. Longitudinal estimates require that data records for each person be linked across interviews, where as cross-sectional estimates do not. For example, annual income estimates obtained by summing the 12 monthly income amounts for each person would require linking records and so would be longitudinal estimates. Because there is no linkage between interviews, cross-sectional estimates can combine data from different interviews only at the aggregate level. Longitudinal person weights were developed for longitudinal estimation, but may be used for cross-sectional estimation as well. However, note that wave files with cross-sectional weights are also produced for the SIPP. Because of the larger sample size with positive weights available on the wave files, it is recommended that these files be used for cross-sectional estimation, if possible.

In this section, it is assumed that all four rotation groups are used for estimation.

Some basic types of longitudinal and cross-sectional estimates which can be constructed using longitudinal person weights are described below in terms of estimated numbers. Of course, more complex estimates, such as percentages, averages, ratios, etc., can be constructed from the estimated numbers. Longitudinal person weights can be used to construct the following types of longitudinal estimates:

1. The number of people who have ever experienced a characteristic during a given time period.

   To construct such an estimate, use the longitudinal person weight for the shortest time period which covers the entire time period of interest. Then, sum the weights over all people who possessed the characteristic of interest at some point during the time period of interest. For example, to estimate the number of people who ever received food stamps during the last six months of 2004, use the CY01 weights, since CY01 weights cover all 12 months of 2004. The same estimate could be generated using the Panel weights, but there may be fewer positively weighted people than in the calendar year.

2. The amount of a characteristic accumulated by people during a given time period.

   To construct such an estimate, use the longitudinal person weight for the shortest time period which covers the entire time period of interest. Then compute the product of the weight times the amount of the characteristic and sum this product over all appropriate people. For example, to estimate the aggregate 2004 annual income of people who were employed during all 12 months of the year, use the CY01 weights. The same estimate could be generated using the Panel weights.

3. The average number of consecutive months of possession of a characteristic (i.e., the average spell length for a characteristic) during a given time period.

   For example, one could estimate the average length of each spell of receiving food stamps during 2004. Also, one could estimate the average spell of unemployment that elapsed before a person found a new job. To construct such an estimate, first identify the people who possessed the characteristic at some point during the time period of interest. Then, create two sums of these person's appropriate longitudinal weights: (1) sum the product of the weight times the number of months the spell lasted and (2) sum the weights only. Now, the estimated average spell length in months is given by (1) divided by (2). A person who experienced two spells during the time period of interest would be treated as two people and appear twice in sums (1) and (2). An alternate method of calculating the average can be found in the section "Standard Error of a Mean or Aggregate."

4. The number of month-to-month changes in the status of a characteristic (i.e., number of transitions) summed over every set of two consecutive months during the time period of interest. To construct such an estimate, sum the appropriate longitudinal person weight each time a change is reported between two consecutive months during the time period of interest. For example, to estimate the number of people who changed from receiving food stamps in July 2004 to not receiving in August 2004, add together the CY01 weights of each person who had such a change. To estimate the number of changes in monthly salary income during the third quarter of 2004, sum together the estimate of number of people who made a change between July 1 and August 1, between August 1 and September 1, and between September 1 and October 1.

   Note that spell and transition estimates should be used with caution because of the biases that are associated with them. Sample people tend to report the same status of a characteristic for all four months of a reference period. This tendency results in a bias toward reported spell lengths that are multiples of four months. This tendency also affects transition estimates in that, for many characteristics, the number of characteristics, the number of month-to-month transitions reported between the last month of one reference period and the first month of the next reference period are much greater than the number of reported transitions between any two months within a reference period. Additionally, spells extending before or after the time period of interest are cut off (censored) at the boundaries of the time period. If they are used in estimating average spell length, a downward bias will result.

   Also using longitudinal person weights one can construct the following type of cross-sectional estimate:

5. Monthly estimates of a characteristic averaged over a number of consecutive months.

   For example, one could estimate the monthly average number of food stamp recipients over the months July through December 2004. To construct such an estimate, first form an estimate for each month in the time period of interest. Use the longitudinal person weight, summing over all people who possessed the characteristic of interest during the month of interest. Then, sum the monthly estimates and divide by the number of months. Either the CY01 or the Panel weights can be used for this calculation.

Adjusting Estimates Which use Less than the Full Sample. When estimates for months with less than four rotations worth of data are constructed from a wave file, factors greater than 1 must be applied. However, when core data from consecutive waves are used together, data from all four rotations may be available, in which case the factors are equal to 1.

All four rotation groups of data are not available for reference months October 2003 through December 2003 and October 2007 through December 2007 (see Table 1). If the time period of interest for a given estimate (of person or household characteristics) includes these months, the estimate may need to be adjusted in some way to account for the missing rotation groups. For longitudinal estimates (types 1-4), this adjustment factor equals four divided by the number of rotation groups contributing data. For example, if the time period of interest for a given estimate is October 2003, then data will be available only from rotation group 1. Therefore, a factor of 4/1 = 4.0000 will be applied. To estimate the number of people ever unemployed in the fourth quarter of 2007, only data from six months is available. Thus, a factor of 1.8519 will be applied. See Table 3 for more information.

Note that, if the given estimate is an average of monthly estimates (estimate type 5), then the number of rotation groups and the factor used will be determined independently for each month in the average and the adjusted monthly estimates will be averaged together in the usual way. For example, to estimate the average number of people unemployed per month in the fourth quarter of 2007, the October, November, and December data will be multiplied by 4/3, 4/2, and 4/1 respectively before being summed together and divided by three.

ACCURACY OF ESTIMATES

SIPP estimates are based on a sample; they may differ somewhat from the figures that would have been obtained if a complete census had been taken using the same questionnaire, instructions, and enumerators. There are two types of errors possible in an estimate based on a sample survey: sampling and nonsampling. We are able to provide estimates of the magnitude of SIPP sampling error, but this is not true of nonsampling error.

Nonsampling Error. Nonsampling errors can be attributed to many sources:
- inability to obtain information about all cases in the sample
- definitional difficulties
- differences in the interpretation of questions
- inability or unwillingness on the part of the respondents to provide correct information
- errors made in the following: collection such as in recording or coding the data, processing the data, estimating values for missing data
- biases resulting from the differing recall periods caused by the interviewing pattern used and undercoverage.

Quality control and edit procedures were used to reduce errors made by respondents, coders and interviewers. More detailed discussions of the existence and control of nonsampling errors in the SIPP can be found in the *SIPP Quality Profile, 1998 SIPP Working Paper Number 230, issued May 1999*.

Undercoverage in SIPP results from missed HUs and missed persons within sample HUs. It is known that undercoverage varies with age, race, and sex. Generally, undercoverage is larger for males than for females and larger for Blacks than for non-Blacks. Ratio estimation to independent age-race-sex population controls partially corrects for the bias due to survey undercoverage. However, biases exist in the estimates to the extent that persons in missed households or missed persons in interviewed households have characteristics different from those of interviewed persons in the same age-race-sex group.

A common measure of survey coverage is the coverage ratio, the estimated population before ratio adjustment divided by the independent population control. Table A below shows SIPP coverage ratios for age-sex-race groups for the Panel weights prior to the ratio adjustment. The SIPP coverage ratios exhibit some variability from month to month, but these are a typical set of coverage ratios. Other Census Bureau household surveys [like the Current Population Survey] experience similar coverage.

Comparability with Other Estimates. Caution should be exercised when comparing this data with data from other SIPP products or with data from other surveys. The comparability problems are caused by such sources as the seasonal patterns for many characteristics, different nonsampling errors, and different concepts and procedures. Refer to the *SIPP Quality Profile* for known differences with data from other sources and further discussions.

Sampling Variability. Standard errors indicate the magnitude of the sampling error. They also partially measure the effect of some nonsampling errors in response and enumeration, but do not measure any systematic biases in the data. The standard errors for the most part measure the variations that occurred by chance because a sample rather than the entire population was surveyed.

Table A. SIPP Average Coverage Ratios for the Panel Weights by Age, Race and Sex

	White Only		*Black Only*		*Residual*	
Age	***Male***	***Female***	***Male***	***Female***	***Male***	***Female***
0-4	0.910	0.900	0.839	0.753	1.205	1.125
5-9	0.927	0.923	0.837	0.780	1.163	1.092
10-14	0.900	0.909	0.830	0.878	1.112	1.069
15-24	0.748	0.758	0.692	0.720	0.902	0.905
25-34	0.784	0.884	0.666	0.788	0.933	0.964
35-44	0.890	0.899	0.802	0.845	1.029	1.016
45-54	0.872	0.926	0.797	0.925	1.034	1.053
55-64	0.918	0.974	0.909	1.004	1.030	1.194
65+	0.995	0.968	0.978	1.159	1.013	1.020

USES AND COMPUTATION OF STANDARD ERRORS

Confidence Intervals. The sample estimate and its standard error enable one to construct confidence intervals, ranges that would include the average result of all possible samples with a known probability. For example, if all possible samples were selected, each of these being surveyed under essentially the same conditions and using the same sample design, and if an estimate and its standard error were calculated from each sample, then:

1. Approximately 68 percent of the intervals from one standard error below the estimate to one standard error above the estimate would include the average result of all possible samples.

2. Approximately 90 percent of the intervals from 1.6 standard errors below the estimate to 1.6 standard errors above the estimate would include the average result of all possible samples.

3. Approximately 95 percent of the intervals from two standard errors below the estimate to two standard errors above the estimate would include the average result of all possible samples.

The average estimate derived from all possible samples is or is not contained in any particular computed interval. However, for a particular sample, one can say with a specified confidence that the average estimate derived from all possible samples is included in the confidence interval.

Hypothesis Testing. Standard errors may also be used for hypothesis testing, a procedure for distinguishing between population characteristics using sample estimates. The most common types of hypotheses tested are 1) the population characteristics are identical versus 2) they are different. Tests may be performed at various levels of significance, where a level of significance is the probability of concluding that the characteristics are different when, in fact, they are identical.
To perform the most common test, compute the difference $X_A - X_B$, where X_A and X_B are sample estimates of the characteristics of interest. A later section explains how to derive an estimate of the standard error of the difference $X_A - X_B$. Let that standard error be S_{DIFF}. If $X_A - X_B$ is between -1.6 times S_{DIFF} and +1.6 times S_{DIFF}, no conclusion about the characteristics is justified at the 10 percent significance level. If, on the other hand, $X_A - X_B$ is smaller than -1.6 times S_{DIFF} or larger than +1.6 times S_{DIFF}, the observed difference is significant at the 10 percent level. In this event, it is commonly accepted practice to say that the characteristics are different. We recommend that users report only those differences that are significant at the 10 percent level or better. Of course, sometimes this conclusion will be wrong. When the characteristics are the same, there is a 10 percent chance of concluding that they are different.

Note that as more tests are performed, more erroneous significant differences will occur. For example, at the 10 percent significance level, if 100 independent hypothesis tests are performed in which there are no real differences, it is likely that about 10 erroneous differences will occur. Therefore, the significance of any single test should be interpreted cautiously. A Bonferroni correction can be done to account for this potential problem that consists of dividing your stated level of confidence by the number of tests you are performing. This correction results in a conservative test of significance.

Note Concerning Small Estimates and Small Differences. Because of the large standard errors involved, there is little chance that estimates will reveal useful information when computed on a base smaller than 75,000. Also, nonsampling error in one or more of the small number of cases providing the estimate can cause large relative error in that particular estimate. Care must be taken in the interpretation of small differences since even a small amount of nonsampling error can cause a borderline difference to appear significant or not, thus distorting a seemingly valid hypothesis test.

Calculating Standard Errors for SIPP Estimates. There are three main ways we calculate the Standard Errors (SEs) for SIPP Estimates. They are as follows:

- Direct estimates using replicate weighting methods;
- Generalized variance function parameters (denoted as *a* and *b*); and
- Simplified tables of SEs based on the *a* and *b* parameters.

While the replicate weight methods provide the most accurate variance estimates, this approach requires more computing resources and more expertise on the part of the user. The Generalized Variance Function (GVF) parameters provide a method of balancing accuracy with resource usage as well as smoothing effect on SE estimates across time. SIPP uses the Replicate Weighting Method to produce GVF parameters(see K. Wolter, *Introducation to Variance Estimation*, Chapter 5 for more information). The GVF parameters are used to create the simplified tables of SEs.

Standard Error Parameters and Tables and Their Use. Most SIPP estimates have greater standard errors than those obtained through a simple random sample because of its two stage cluster sample design. To derive standard errors that would be applicable to a wide variety of estimates and could be prepared at a moderate cost, a number of approximations were required. Estimates with similar standard error behavior were grouped together and two parameters (denoted a and b) were developed to approximate the standard error behavior of each group of estimates. Because the actual standard error behavior was not identical for all estimates within a group, the standard errors computed from these parameters provide an indication of the order of magnitude of the standard error for any specific estimate. These a and b parameters vary by characteristic and by demographic subgroup to which the estimate applies. Table 2 provides base a and b parameters to be used for the 2004 Panel estimates created using either the calendar year and the 4-Wave Panel weights.

In this section we discuss the adjustment of base "a" and "b" parameters to provide "a" and "b" parameters appropriate for each type of longitudinal and cross-sectional estimate described in the section "Use of Person Weights." Later sections will discuss the use of the adjusted parameters in various formulas to compute standard errors of estimated numbers, percents, averages, etc. Table 2 provide the base "a" and "b" parameters needed to compute the approximate standard errors for estimates using 4-Wave panel or calender year weights. Table 3 provides additional factors to be used for averages of monthly cross-sectional estimates. These factors are needed for two reasons: the monthly estimates are correlated and averaging over a greater number of monthly estimates will produce an average with a smaller standard error. Table 5 gives correlations between quarterly and yearly averages of cross-sectional estimates. These correlations are used in the formula for the standard error of a difference (Formula (9)).

The creation of appropriate "a" and "b" parameters for the previously discussed types of estimates are described below. Again, it is assumed that all four rotation groups are used in estimation. If not, refer to the section "Adjusting Standard Errors of Estimates Which Use Less Than the Full Sample."

1. The number of people who have ever experienced a characteristic during a given time period.

   The appropriate "a" and "b" parameters are taken directly from Table 2. The choice of parameter depends on the weights used, on the characteristic of interest, and on the demographic subgroup of interest.

2. Amount of a characteristic accumulated by people during a given time period.

   The appropriate "b" parameters are also taken directly from Table 2.

3. The average number of consecutive months of possession of a characteristic per spell (i.e., the average spell length for a characteristic) during a given time period.

   Start with the appropriate base "a" and "b" parameters from Table 2. The parameters are then inflated by an additional factor, g, to account for people who experience multiple spells during the time period of interest. This factor is computed by:

$$g = \frac{\sum_{i=1}^{n} m_i^2}{\sum_{i=1}^{n} m_i}, \qquad (1)$$

   where there are n people with at least one spell and m_i is the number of spells experienced by person I during the time period of interest.

4. The number of month-to-month changes in the status of a characteristic (i.e., number of transitions) summed over every set of two consecutive months during the time period of interest.

   Obtain a set of adjusted "a" and "b" parameters exactly as just described in 3, then multiply these parameters by an additional factor. Use 1.0000 if the time period of interest is two months and 2.0000 for a longer time period. (The factor of 2.0000 is based on the conservative assumption that each spell produces two transitions within the time period of interest.)

5. Monthly estimates of a characteristic averaged over a number of consecutive months.

Appropriate base "a" and "b" parameters are taken from Table 2. If more than one longitudinal weight has been used in the monthly average, then there is a choice of parameters from Table 2. Choose the table which gives the largest parameter. Next multiply the base "a" and "b" parameters by the factor from Table 3 corresponding to the number of months in the average.

Adjusting Standard Error Parameters for Estimates which Use Less Than the Full Sample. If some rotation groups are unavailable to contribute data to a given estimate, then the estimate and its standard error need to be adjusted. The adjustment of the estimate is described in a previous section. The standard error of a longitudinal estimates (type 1-4) is adjusted by multiplying the appropriate “a” and “b” parameters by a factor equal to four divided by the number of rotation groups contributing data to the estimate. Note that the parameters for the standard error of an average must still be adjusted according to this rule, even though the average itself is unaffected by the adjustment for missing rotation groups.

For the standard error of cross-sectional estimates which cover only one month, the factor can be computed as just described or it can be taken from Table 3 where the factor is given for each single reference month, October 2003 to March 2007. For the standard error of quarterly averages of month estimates which use less than the full sample, special factors are used, also given in Table 3 for the fourth quarter of 2003 to the fourth quarter of 2007.

Standard Errors of Estimated Numbers. The approximate standard error, s_x, of an estimated number of people may be obtained by using the formula:

$$s_x = \sqrt{ax^2 + bx} \qquad (2)$$

Here x is the size of the estimate and a and b are the parameters associated with the particular type of characteristic being estimated. Note that this method should not be applied to dollar values.

<u>Illustration.</u>

Suppose the SIPP estimate of the number of people ever receiving Social Security during the first three months of 2004 is 38,122,000. (This estimate is obtained using the 2004 Calendar year weight.) The appropriate "a" and "b" parameters to use in calculating a standard error for the estimate are obtained from Table 2. They are a = -0.00002130, b = 4,820, respectively. Using Formula (2), the approximate standard error is

$$\sqrt{(-0.00002130)(38{,}122{,}000)^2 + (4{,}820)(38{,}122{,}000)} = 390{,}887 \ \textit{persons}$$

The 90-percent confidence interval as shown by the data is from 37,478,990 to 38,765,010.

Therefore, a conclusion that the average estimate derived from all possible samples lies within a range computed in this way would be correct for roughly 90 percent of all samples. Similarly, the 95-percent confidence interval as shown by the data is from 37,355,860 to 38,888,139 and we could conclude that the average estimate derived from all possible samples lies within this interval.

Standard Error of a Mean. A mean is defined here to be the average quantity of some item (other than people, families, or households) per person. For example, it could be the annual household income of females age 25 to 34. The standard error of a mean can be approximated by Formula (3) below. Because of the approximations used in developing Formula (3), an estimate of the standard error of the mean obtained from this formula will generally underestimate the true standard error. The formula used to estimate the standard error of a mean $\bar{x}$ is

$$s_{\bar{x}} = \sqrt{\left(\frac{b}{y}\right)s^2} \qquad (3)$$

where y is the size of the base, s^2 is the estimated population variance of the item and b is the parameter associated with the particular type of item.

The population variance s^2 may be estimated by one of two methods. In both methods, we assume x_i is the value of the item for unit "I." (Unit may be person, family, or household). To use the first method, the range of values for the item is divided into "c" intervals. The upper and lower boundaries of interval j are Z_{j-1} and Z_j, respectively. Each unit is placed into one of "c" groups such that $Z_{j-1} < x_i < Z_j$.

The estimated population mean, $\bar{x}$, and variance, s^2, are given by the formulas:

$$\bar{x} = \sum_{j=1}^{c} p_j m_j$$
$$s^2 = \sum_{j=1}^{c} p_j m_j^2 - \bar{x}^2, \qquad (4)$$

where p_j is the estimated proportion of units in group j , and $m_j = (Z_{j-1} + Z_j)/2$. The most representative value of the item in group j is assumed to be m_j. If group "c" is open-ended, or there exists no upper interval boundary, then an approximate value for m_c is

$$m_c = \frac{3}{2} Z_{c-1} .$$

In the second method, the estimated population mean, $\bar{x}$, and variance, s^2, are given by the formulas

$$\bar{x} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}$$

$$s^2 = \frac{\sum_{i=1}^{n} w_i x_i^2}{\sum_{i=1}^{n} w_i} - \bar{x}^2, \qquad (5)$$

where there are n units with the item of interest and w_i is the final weight for unit "I" (note that $\sum w_i = y$).

Illustration of Method 1.

Suppose that the 2004 distribution of annual incomes is given in Table 4 for people aged 25 to 34 who were employed for all 12 months of 2004.

$$\bar{x} = \frac{370}{23{,}527}(2{,}500) + \frac{302}{23{,}527}(6{,}250) + \ldots + \frac{2{,}138}{23{,}527}(105{,}000) = \$38{,}704.$$

The mean annual cash income from following formula is

Using Formula (4) and the mean annual cash income of $38,704 the estimated population variance, s^2, is

$$s^2 = \frac{370}{23{,}527}(2{,}500)^2 + \frac{302}{23{,}527}(6{,}250)^2 + \ldots + \frac{2{,}138}{23{,}527}(105{,}000)^2 - (38{,}704)^2 = 649{,}457{,}303.$$

The appropriate "b" parameter from Table 2 is 4,280. Now, using Formula (3), the estimated standard error of the mean is

$$s_{\bar{x}} = \sqrt{\frac{4{,}820}{23{,}527{,}377}(649{,}457{,}303)} = \$365$$

Illustration of Method 2.

Suppose that we are interested in estimating the average length of spells of food stamp recipiency during the calendar year 2004 for a given subpopulation. Also, suppose there are only 10 sample people in the subpopulation who were food stamp recipients. (This example is a hypothetical situation used for illustrative purposes only; actually, 10 sample cases would be too few for a reliable estimate and their weights could be substantially different from those given.) The number of consecutive months of food stamp recipiency during 2004 and the CY1 weights are given below for each sample person:

Sample Person	Spell Length (in months)	CY01 Weight
1	4,3	5,300
2	5	7,100
3	9	4,900
4	3,3,2	6,500
5	12	9,200
6	12	5,900
7	4,1	7,600
8	7	4,200
9	6	5,500
10	4	5,700

Using the following formula , the average spell of food stamp recipiency is estimated to be

$$\bar{x} = \frac{(5300)(4) + (5300)(3) + \ldots + (5700)(4)}{5300 + 5300 + \ldots + 5700}$$

$$= 473{,}100/87{,}800$$

$$= 5.4 \ \textit{months}$$

The standard error will be computed by Formula (3). First, the estimated population variance can be obtained by Formula (5):

$$s^2 = \frac{(5300)(4)^2 + (5300)(3)^2 + \ldots + (5700)(4)^2}{5300 + 5300 + \ldots + 5700} - (5.4)^2$$

$$= 12.4 \ (\textit{months})^2$$

$$s_x = \sqrt{(b)(y)s^2} \qquad (6)$$

Next, the base "b" parameter of 4,463 is taken from Table 2 and multiplied by the factor computed from Formula (1):

$$g = \frac{2^2 + 1 + 1 + 3^2 + 1 + 1 + 2^2 + 1 + 1 + 1}{2 + 1 + 1 + 3 + 1 + 1 + 2 + 1 + 1 + 1}$$

$$= 1.71$$

Therefore, the final "b" parameter is 7,632, and the standard error of the mean is:

$$s = \sqrt{\frac{7{,}632}{87{,}800} (12.4)} = 1.04 \ \textit{months}$$

Standard error of an Aggregate. An aggregate is defined to be the total quantity of an item summed over all the units in a group. The standard error of an aggregate can be approximated using Formula (6).

As with the estimate of the standard error of a mean, the estimate of the standard error of an aggregate will generally underestimate the true standard error. Let y be the size of the base, s^2 be the estimated population variance of the item obtained using Formula (4) or Formula (5) and b be the parameter associated with the particular type of item. The standard error of an aggregate is:

$$s_x = \sqrt{(b)\ (y)s^2} \qquad (6)$$

Standard Errors of Estimated Percentages. The reliability of an estimated percentage, computed using sample data for both numerator and denominator, depends upon both the size of the percentage and the size of the total upon which the percentage is based. Estimated percentages are relatively more reliable than the corresponding estimates of the numerators of the percentages, particularly if the percentages are 50 percent or more, e.g., the percent of people employed is more reliable than the estimated number of people employed. When the numerator and denominator of the percentage have different parameters, use the parameter (and appropriate factor) of the numerator. If proportions are presented instead of percentages, note that the standard error of a proportion is equal to the standard error of the corresponding percentage divided by 100.

There are two types of percentages commonly estimated. The first is the percentage of people sharing a particular characteristic such as the percent of people owning their own home. The second type is the percentage of money or some similar concept held by a particular group of people or held in a particular form. Examples are the percent of total wealth held by people with high income and the percent of total income received by people on welfare.

For the percentage of people, the approximate standard error, $s_{(x,p)}$, of the estimated percentage p may be approximated by the formula

$$s_{(x,p)} = \sqrt{\frac{b}{x}\,(p)(100-p)}. \tag{7}$$

Here x is the size of the subclass of social units which is the base of the percentage, p is the percentage $(0 < p < 100)$, and b is the parameter associated with the characteristic in the numerator.

<u>Illustration.</u>
Suppose that using the 4-Wave weight, it was estimated that 59,355,000 males were employed in July 2004 and an estimated 2.4 percent of them became unemployed in August 2004. The base "b" parameter is 4,690 (from Table 2). Using Formula (7) and the appropriate "b" parameter, the approximate standard error is

$$\sqrt{\frac{(4{,}690)}{(59{,}355{,}000)}\,(2.4)(100-2.4)} = 0.14 \textit{ percent}.$$

Consequently, the 90-percent confidence interval as shown by these data is from 2.18 to 2.62 percent.

For percentages of money, a more complicated formula is required. A percentage of money will usually be estimated in one of two ways. It may be the ratio of two aggregates:

$$p_I = 100\ (x_A\ /\ x_N)$$

or it may be the ratio of two means with an adjustment for different bases:

$$p_I = 100\ (\hat{p}_A\ \bar{x}_A\ /\ \bar{x}_N)$$

where x_A and x_N are aggregate money figures, $\bar{x}_A$ and $\bar{x}_N$ are mean money figures, and $\hat{p}_A$ is the estimated number in group A divided by the estimated number in group N.

In either case, we estimate the standard error as

$$s_I = \sqrt{\left(\frac{\hat{p}_A \bar{x}_A}{\bar{x}_N}\right)^2 \left[\left(\frac{s_p}{\hat{p}_A}\right)^2 + \left(\frac{s_A}{\bar{x}_A}\right)^2 + \left(\frac{s_B}{\bar{x}_N}\right)^2\right]} \,, \tag{8}$$

where s_p is the standard error of $\hat{p}_A$, s_A is the standard error of $\bar{x}_A$ and s_B is the standard error of $\bar{x}_N$. To calculate s_p, use Formula (7). The standard errors of $\bar{x}_N$ and $\bar{x}_A$ may be calculated using Formula (3).

It should be noted that there is frequently some correlation between $\hat{p}_A$, $\bar{x}_N$ and $\bar{x}_A$. Depending on the magnitude and sign of the correlations, the standard error will be over or underestimated.

Illustration.
Suppose that in October 2004 an estimated 8.8% of males 16 years and over were black, the mean monthly earnings of these black males was $1288, the mean monthly earnings of all males 16 years and over was $1911, and the corresponding standard errors are .28%, $36, and $27. Then, the percent of male earnings made by blacks in October 2004 is:

$$p_M = .088 \left(\frac{1288}{1911}\right) x \ 100$$

$$= 5.9\%$$

Using Formula (8), the approximate standard error is:

$$s_M = \sqrt{\left(\frac{(.088)(1288)}{1911}\right)^2 \left[\left(\frac{.0028}{.0880}\right)^2 + \left(\frac{36}{1288}\right)^2 + \left(\frac{27}{1911}\right)^2\right]}$$

$$= 0.26\%$$

Standard Error of a Difference. The standard error of a difference between two sample estimates is approximately equal to

$$s_{(x-y)} = \sqrt{s_x^{\,2} + s_y^{\,2} - r\, s_x\, s_y} \,, \tag{9}$$

where s_x and s_y are the standard errors of the estimates x and y.

The estimates can be numbers, percents, ratios, etc. The correlation between x and y is represented by r. Some correlations are given in Table 5. The above formula assumes that the correlation coefficient between the characteristics estimated by x and y is non-zero. If no correlations has been provided for a given set of x and y estimates, assume r = 0. However, if the correlation is really positive (negative), then this assumption will tend to cause overestimates (underestimates) of the true standard error.

Illustration.
Suppose that SIPP estimates show the number of people age 35-44 years with annual cash income of $50,000 to $59,999 was 3,186,000 in 2004 and the number of people age 25-34 years with annual cash income of $50,000 to $59,999 in the same time period was 2,619,000. Then, using parameters from Table 2 and Formula (2), the standard errors of these numbers are approximately 121,319 and 110,144, respectively. The difference in sample estimates is 567,000 and using Formula (9), the approximate standard error of the difference is

$$\sqrt{(121{,}319)^2 + (110{,}144)^2} = 163{,}860 \ .$$

Suppose that it is desired to test at the 10 percent significance level whether the number of people with annual cash income of $50,000 to $59,999 was different for people age 35-44 years than for people age 25-34 years. To perform the test, compare the difference of 567,000 to the product $1.6 * 216{,}471 = 269{,}549$. Since the difference is larger than 1.6 times the standard error of the difference, the data show that the two age groups are significantly different at the 10 percent significance level

Standard Error of a Median. The median quantity of some item such as income for a given group of people is that quantity such that at least half the group have as much or more and at least half the group have as much or less. The sampling variability of an estimated median depends upon the form of the distribution of the item as well as the size of the group. To calculate standard errors on medians, the procedure described below may be used.

The median, like the mean, can be estimated using either data which has been grouped into intervals or ungrouped data. If grouped data are used, the median is estimated using Formulas (10) or (11) with p = 0.5. If ungrouped data are used, the data records are ordered based on the value of the characteristic, then the estimated median is the value of the characteristic such that the weighted estimate of 50 percent of the subpopulation falls at or below that value and 50 percent is at or above that value. Note that the method of standard error computation which is presented here requires the use of grouped data. Therefore, it should be easier to compute the median by grouping the data and using Formulas (10) or (11).

An approximate method for measuring the reliability of an estimated median is to determine a confidence interval about it. (See the section on sampling variability for a general discussion of confidence intervals.) The following procedure may be used to estimate the 68-percent confidence limits and hence the standard error of a median based on sample data.

- Determine, using Formula (7), the standard error of an estimate of 50 percent of the group.

- Add to and subtract from 50 percent the standard error determined in step 1.

- Using the distribution of the item within the group, calculate the quantity of the item such that the percent of the group with more of the item is equal to the smaller percentage found in step 2. This quantity will be the upper limit for the 68-percent confidence interval. In a similar fashion, calculate the quantity of the item such that the percent of the group with more of the item is equal to the larger percentage found in step 2. This quantity will be the lower limit for the 68-percent confidence interval.

- Divide the difference between the two quantities determined in step 3 by two to obtain the standard error of the median.

To perform step 3, it will be necessary to interpolate. Different methods of interpolation may be used. The most common are simple linear interpolation and Pareto interpolation. The appropriateness of the method depends on the form of the distribution around the median. If density is declining in the area, then we recommend Pareto interpolation. If density is fairly constant in the area, then we recommend linear interpolation. Note, however, that Pareto interpolation can never be used if the interval contains zero or negative measures of the item of interest. Interpolation is used as follows. The quantity of the item such that p percent have more of the item is

$$X_{pN} = exp\left[\left(Ln\left(\frac{pN}{N_1}\right) \bigg/ Ln\left(\frac{N_2}{N_1}\right)\right) \quad Ln\left(\frac{A_2}{A_1}\right)\right] A_1 \ . \qquad (10)$$

if Pareto Interpolation is indicated and

$$X_{pN} = \left[\frac{pN-N_1}{N_2-N_1} \quad (A_2-A_1) + A_1\right] \qquad (11)$$

if linear interpolation is indicated, where

N	is the size of the group,
A_1 and A_2	are the lower and upper bounds, respectively, of the interval in which X_{pN} falls,
N_1 and N_2	are the estimated number of group members owning more than A_1 and A_2, respectively,
exp	refers to the exponential function and
Ln	refers to the natural logarithm function

Illustration.
To illustrate the calculations for the sampling error on a median, we return to Table 4. The median annual income for this group is $32,200. The size of the group is 23,527,000.

1. Using Formula (7), the standard error of 50 percent on a base of 23,527,000 is about 0.71 percentage points.

2. Following step 2, the two percentages of interest are 49.29 and 50.71.

3. By examining Table 4, we see that the percentage 49.29 falls in the income interval from 30,000 to 39,999. (Since 54.7% receive more than $30,000 per month, the dollar value corresponding to 49.29 must be between $30,000 and $39,999). Thus, $A_1 = \$30{,}000$, $A_2 = \$39{,}999$, $N_1 = 18{,}377{,}000$, and $N_2 = 12{,}881{,}000$.

In this case, we decided to use Pareto interpolation. Therefore, the upper bound of a 68% confidence interval for the median is

$$\$30{,}000 \ \exp\left[\left(Ln\left(\frac{(.4929)(23{,}527{,}000)}{18{,}377{,}000}\right) \Big/ Ln\left(\frac{12{,}881{,}000}{18{,}377{,}000}\right)\right) Ln\left(\frac{39{,}999}{30{,}000}\right)\right] = \$43{,}549 \ .$$

Also by examining Table 4, we see that 50.71 falls in the same income interval. Thus, A_1, A_2, N_1, and N_2 are the same. We also use Pareto interpolation for this case. So the lower bound of a 68% confidence interval for the median is

$$\$30{,}000 \ \exp\left[\left(Ln\left(\frac{(.5071)(23{,}527{,}000)}{18{,}377{,}000}\right) \Big/ Ln\left(\frac{12{,}881{,}000}{18{,}377{,}000}\right)\right) Ln\left(\frac{39{,}999}{30{,}000}\right)\right] = \$42{,}560 \ .$$

Thus, the 68-percent confidence interval on the estimated median is from $42,560 to $43,549.

An approximate standard error is

$$\frac{\$43,549 \ - \ \$42,560}{2} \ = \ \$494.50 \ .$$

Standard Errors of Ratios of Means and Medians. The standard error for a ratio of means or medians is approximated by:

$$s_{\left(\frac{x}{y}\right)} = \sqrt{\left(\frac{x}{y}\right)^2 \left[\left(\frac{s_y}{y}\right)^2 + \left(\frac{s_x}{x}\right)^2\right]} \tag{12}$$

where x and y are the means or medians, and s_x and s_y are their associated standard errors. Formula (12) assumes that the means are not correlated. If the correlation between the population means estimated by x and y are actually positive (negative), then this procedure will tend to produce overestimates (underestimates) of the true standard error for the ratio of means.

Standard Errors Using SAS or SPSS. Standard errors and their associated variance, calculated by SAS or SPSS statistical software package, do not accurately reflect the SIPP's complex sample design. Erroneous conclusions will result if these standard errors are used directly. We provide adjustment factors by characteristics that should be used to correctly compensate for likely under-estimates. The factors called DEFF available in Table 2, must be applied to SAS or SPSS generated variances. The square root of DEFF can be directly applied to similarly generated standard errors. These factors approximate design effects which adjust statistical measures for sample designs more complex than simple random sample.

Table 1 - SIPP Panel 2004 Reference Months (horizontal) for Each Interview Month (vertical)

Month of Interview	Wave / Rotation	2003			2004												2005												2006												2007											
		4th Quarter			1St Quarter			2nd Quarter			3rd Quarter			4Th Quarter			1St Quarter			2nd Quarter			3rd Quarter			4Th Quarter			1St Quarter			2nd Quarter			3rd Quarter			4Th Quarter			1St Quarter			2nd Quarter			3rd Quarter			4Th Quarter		
		Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Feb 04	1/1	1	2	3	4																																															
Mar	1/2		1	2	3	4																																														
Apr	1/3			1	2	3	4																																													
May	1/4				1	2	3	4																																												
Jun	2/1					1	2	3	4																																											
July	2/2						1	2	3	4																																										
Aug	2/3							1	2	3	4																																									
Sept	2/4								1	2	3	4																																								
Oct	3/1									1	2	3	4																																							
Nov	3/2										1	2	3	4																																						
Dec	3/3											1	2	3	4																																					
Jan 05	3/4												1	2	3	4																																				
Feb	4/1													1	2	3	4																																			
Mar	4/2														1	2	3	4																																		
Apr	4/3															1	2	3	4																																	
May	4/4																1	2	3	4																																
Jun	5/1																	1	2	3	4																															
July	5/2																		1	2	3	4																														
Aug	5/3																			1	2	3	4																													
Sept	5/4																				1	2	3	4																												
Oct	6/1																					1	2	3	4																											
Nov	6/2																						1	2	3	4																										
Dec	6/3																							1	2	3	4																									
Jan 06	6/4																								1	2	3	4																								
Feb	7/1																									1	2	3	4																							
Mar	7/2																										1	2	3	4																						
Apr	7/3																											1	2	3	4																					
May	7/4																												1	2	3	4																				
Jun	8/1																													1	2	3	4																			
July	8/2																														1	2	3	4																		
Aug	8/3																															1	2	3	4																	
Sep	8/4																																1	2	3	4																
Oct	9/1																																	1	2	3	4															
Nov	9/2																																		1	2	3	4														
Dec	9/3																																			1	2	3	4													
Jan 07	9/4																																				1	2	3	4												
Feb	10/1																																					1	2	3	4											
Mar	10/2																																						1	2	3	4										
Apr	10/3																																							1	2	3	4									
May	10/4																																								1	2	3	4								
Jun	11/1																																									1	2	3	4							
Jul	11/2																																										1	2	3	4						
Aug	11/3																																											1	2	3	4					
Sep	11/4																																												1	2	3	4				
Oct	12/1																																													1	2	3	4			
Nov	12/2																																														1	2	3	4		
Dec	12/3																																															1	2	3	4	
Jan 08	12/4																																																1	2	3	4

Table 2: SIPP Generalized Variance Parameters for Calendar Year 2004

Characteristics	**Parameters**		
Individuals	***a***	***b***	**DEFF**
Poverty and Program Participation	-0.00001972	4463	1.82
Male	-0.00004088	4463	
Female	-0.00003811	4463	
Income and Labor Force	-0.00002130	4820	1.96
Male	-0.00004414	4820	
Female	-0.00004116	4820	
Other (Person) Items	-0.00001661	4768	1.94
Male	-0.00003400	4768	
Female	-0.00003249	4768	
Black (Person) Items	-0.00012882	4631	1.88
Male	-0.00027785	4631	
Female	-0.00020402	4631	
Hispanic (Person) Items	-0.00016308	6539	2.66
Male	-0.00031847	6539	
Female	-0.00033423	6539	
Households			
Total or White	-0.00003581	4029	1.64
Black	-0.00027932	4029	
Hispanic	-0.00045451	4029	

Table 3 - Factors to be Applied to Table 2 Base Parameters to Obtain Parameters for Various Reference Periods

# of available rotation months[4] Monthly estimate	Factor
1	4.0000
2	2.0000
3	1.3333
4	1.0000
Quarterly estimate	
6	1.8519
8	1.4074
9	1.2222
10	1.0494
11	1.0370

[4] The number of available rotation months for a given estimate is the sum of the number of rotations available for each month of the estimate.

Table 4 - Hypothetical Distribution of Annual Income Among People 25 to 34 Years Old

Intervals of Annual Cash Income	Total	under $5000	$5000 to $7499	$7500 to $9999	$10000 to $12,499	$12,500 to $14,999	$15,000 to $17,499	$17,500 to $19,999	$20,000 to $29,999	$30,000 to $39,999	$40,000 to $49,999	$50,000 to $59,999	$60,000 to $69,999	$70,000 and over
Mid-intervals of Annual Cash Income		2,500	6250	8750	11,250	13,750	16,250	18,750	25,000	35,000	45,000	55,000	65,000	105,000
Thousands in interval	23,527	370	302	447	685	935	1,113	1,298	5,496	4,596	3,121	1,902	1,124	2,138
Cumulative with at least as much as lower bound of interval		23,527	23,158	22,856	22,409	21,724	20,789	19,675	18,377	12,881	8,285	5,164	3,262	2,138
Percent with at least as much as lower bound of interval		100.0	98.4	97.1	95.2	92.3	88.4	83.6	78.1	54.7	35.2	21.9	13.9	9.1

Table 5 - Correlations between Estimates of the Same Characteristic at Two Points of Time. Both Estimates must be Monthly Estimates Averaged over Quarters or Years

	Quarterly Estimates				
	Consecutive Quarters	1 Quarter Apart	2 Quarters Apart	3 Quarters Apart	Calendar Year Estimates 2001 to 2002
INDIVIDUALS					
A. Both Estimates Created Using The Same Weight, Either 4 Wave, 7 Wave, or 9 Wave Weights					
Income					
Social Security or Private Pensions	0.97	0.86	0.75		
Other	0.72	0.63	0.54		
B. One Estimate Created Using An Annual Weight While The Other Estimate Is Created Using A Different Annual Weight					
Income					
Social Security or Private Pensions	0.81	0.72	0.63	0.55	0.70
Other	0.60	0.53	0.45	0.37	0.49
C.Both Estimates Created Using The 9 Wave (or Panel) Weight					
Income					
Social Security or Private Pensions	0.97	0.86	0.75	0.65	0.83
Other	0.72	0.63	0.54	0.46	0.58